Gaussian Process Modeling of EEG for the Detection of Neonatal Seizures

Stephen Faul*, Gregor Gregorčič, Geraldine Boylan, William Marnane, Gordon Lightbody, and Sean Connolly

***Abstract*—Gaussian process (GP) probabilistic models have attractive advantages over parametric and neural network modeling approaches. They have a small number of tuneable parameters, can be trained on relatively small training sets, and provide a measure of prediction certainty. In this paper, these properties are exploited to develop two methods of highlighting the presence of neonatal seizures from electroencephalograph (EEG) signals. In the first method, the certainty of the GP model prediction is used to indicate the presence of seizures. In the second approach, the hyperparameters of the GP model are used. Tests are carried out with a feature set of ten EEG measures developed from various signal processing techniques. Features are evaluated using a neural network classifier on 51 h of real neonatal EEG. The GP measures, in particular, the prediction certainty approach, produce a high level of performance compared to other modeling methods and methods currently in clinical use for EEG analysis, indicating that they are an important and useful tool for the real-time detection of neonatal seizures.**

***Index Terms*—Electroencephalogram (EEG) modeling, Gaussian process (GP) modeling, neonatal seizure detection.**

I. Introduction

SEIZURES are a common neurological emergency in the Neonatal Intensive Care Unit (NICU). Studies have shown that 0.7 to 2.8 per thousand term newborns develop seizures, and that this rises to 58 to 132 per thousand in preterm infants [1], [2]. Newborns who develop seizures have a high risk of developing long-term neurodevelopmental sequelae; 3% suffer hearing loss, 5% suffer visual loss, and 50% suffer from a delay in psychomotor development or neurological damage [3]. Clearly, the effect on the general quality of life due to neonatal seizures is considerable, and, therefore, their detection and adequate treatment is a high priority.

The main reason why neonatal seizures often remain undetected and, hence, untreated, is that they are very often clinically unsuspected. They may be accompanied by only very subtle signs, such as lip-smacking or ocular fixation, or, in some cases, there may be no clinical signs at all. Consequently, the extent of the electrographic seizure burden in the sick neonate can be greatly underestimated [4]. As effective seizure control in the neonate requires abolition of both clinical and electrographic seizures, electroencephalogram (EEG) monitoring is essential.

Compact digital video-EEG recording systems are now available for use in the NICU. However, particular skills are required to acquire and interpret neonatal EEG as it exhibits a more complex architecture than adult EEG. As many NICUs lack this expertise, they often rely on cerebral function monitors (CFMs) to assess the neurological condition of neonates. Though these systems are easy to use, they can fail to detect many focal or short duration seizures or mistake artifacts for seizure events [5]. Therefore, there is a great need for a reliable method of automatic seizure detection in the NICU. Such a system would provide the best window of opportunity for treatment of neonatal seizures in NICUs that may lack expertise, personnel or time resources. Such a system could also be utilized offline to aid research into the causes and treatments of neonatal seizures.

Previous work in this area, including the use of modeling methods, has proved too unreliable for use in a clinical environment [6]. In this paper, a new approach is taken to the seizure detection problem using Gaussian process (GP) modeling is proposed. Features of the GP model are used to highlight seizure events occurring in the neonatal EEG. It is proposed that these features could be used as part of the feature extraction process of the neonatal seizure detection system outlined in [7].

II. Neonatal EEG and Seizures

The healthy neonatal EEG consists of activity of various frequencies, changing in character with sleep, concentration, movement, and external stimuli. Furthermore, unlike the adult EEG, the neonatal EEG can contain transients such as focal sharp waves, slow waves and, in the preterm neonate, periods of suppression that are very different from activities seen in older children or adults [8]. These differences make the analysis of the neonatal EEG much more complex than that of adult EEG. An example of 38 s of an eight-channel, nonseizure EEG is shown in Fig. 1(a) which shows its varying mixture of varying frequencies.

Neonatal seizure activity generally consists of rhythmic, stereotyped activity and may be focused in one part or spread over large areas of the brain. The neonatal seizure waveform can consist of rhythmic discharges, spikes, sharp waves and slow waves. A recent study has reported a mean seizure duration of 1 min 37 s in preterm infants and 2 min in full-term infants [8]. A 48-s, eight-channel seizure EEG trace is shown in Fig. 1(b), which shows repetitive sharp waves across a number of channels, predominantly in the right central region of the brain.

Manuscript received March 16, 2006; revised February 7, 2007. This work was supported in part by a grant from Enterprise Ireland and in part by an interdisciplinary grant from the HRB. *Asterisk indicates corresponding author.*

*S. Faul is with the Department of Electrical and Electronic Engineering, College Road, University College Cork, Cork, Ireland (e-mail: stephenf@rennes.ucc.ie).

W. Marnane, and G. Lightbody are with the Department of Electrical and Electronic Engineering, University College Cork, Cork, Ireland.

G. Gregorčič is with AVL List GMBH, Hans-List-Platz 1, A-8020 Graz, Austria.

G. Boylan is with the Department of Paediatrics and Child Health, Cork University Hospital, Cork, Ireland.

S. Connolly is with the Department of Clinical Neurophysiology, St. Vincent's Hospital, Dublin, Ireland.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2007.895745

(a)

(b)

Fig. 1. Samples of nonseizure and seizure EEG recordings. Each recording consists of eight channels followed by an electrocardiograph (ECG) trace. The traces are 38 and 48 s in length, respectively. (a) Nonseizure EEG; (b) seizure EEG.

III. Gaussian Process Modeling

Due to the nonlinear nature of the neonatal EEG, nonlinear modeling approaches have been proposed for its analysis in the past, such as fuzzy or neural network models, among others [9], [10]. The difficulties associated with black-box modeling approaches are mainly related to the curse of dimensionality and lack of transparency of the global model. A further problem is the sheer number of tunable parameters of neural network models which need to be trained from a set of training data or provided from prior knowledge.

GP modeling could provide a solution to these problems. The number of tunable parameters for a GP model is greatly reduced over its neural network counterpart. A GP model also provides an estimate of the variance of its predicted output, which can be interpreted as a level of confidence of the model. This measure of variance is a major advantage over neural network or fuzzy models as it gives an indication of when the model can be trusted [11], [12].

A stochastic process is an indexed collection of random variables. It is defined by properties of the joint distribution for this collection. Thus, any finite set of random variables $\underline{y}$, is a GP if it has a joint Gaussian distribution

$$P(\underline{y}\,|\,\boldsymbol{C},\boldsymbol{\Phi}_N)=\frac{1}{Q}e^{-\frac{1}{2}(\underline{y}-\underline{\mu})^T\boldsymbol{C}^{-1}(\underline{y}-\underline{\mu})} \tag{1}$$

where $P(\underline{y}\,|\,\boldsymbol{C},\boldsymbol{\Phi}_N)$ is the distribution of the output $\underline{y}$, given the covariance matrix $\boldsymbol{C}$ (defined using the input matrix Φ_N), Q is the normalizing constant and $\underline{\mu}$ is the mean vector of the distribution. The GP is, therefore, fully represented by its covariance function $C(\cdot)$, which defines $\boldsymbol{C}$, and its mean $\underline{\mu}$. If a zero-mean distribution is assumed, then the process is defined wholly by $\boldsymbol{C}$.

A. Constructing the GP Model

Given a noisy input/output set of data, the full input matrix of N d-dimensional input vectors $\psi(k)$ is constructed by

$$\Phi_N=\begin{bmatrix}\psi_1(1) & \psi_2(1) & \dots & \psi_d(1)\\ \psi_1(2) & \psi_2(2) & \dots & \psi_d(2)\\ \vdots & \vdots & \vdots & \vdots\\ \psi_1(N) & \psi_2(N) & \dots & \psi_d(N)\end{bmatrix}. \tag{2}$$

The corresponding output vector $\underline{y}_N$ is given by

$$\underline{y}_N=[y(1),y(2),\dots,y(N)]^T. \tag{3}$$

The aim is to construct a model from the above data and then for an unseen input vector

$$\psi(N+1)=[\psi_1(N+1),\psi_2(N+1),\dots,\psi_d(N+1)] \tag{4}$$

to find the distribution for the corresponding output $y(N+1)$. This distribution is given by the mean and variance of the output $y(N+1)$ and can be written as

$$P(y(N+1)|D,C(\cdot),\psi(N+1))=\frac{1}{Q}e^{-\frac{1}{2}\left(\underline{y}_{N+1}^T\boldsymbol{C}_{N+1}^{-1}\underline{y}_{N+1}\right)}. \tag{5}$$

The mean and standard deviation of the distribution of $y(N+1)$ given in (5) can be evaluated by inverting $\boldsymbol{C}_{N+1}$. An efficient method of inverting $\boldsymbol{C}_{N+1}$ can be achieved by partitioning the matrix into C_N, $\underline{v}_{N+1}$, $\underline{v}_{N+1}^T$ and ν as shown in Fig. 2(a) and utilizing the partitioned inverse equation [13]. $\boldsymbol{C}_{N+1}^{-1}$ can then be generated as shown in Fig. 2(b), where

$$\tilde{\nu}=(\nu-\underline{v}_{N+1}^T\boldsymbol{C}_N^{-1}\underline{v}_{N+1})^{-1} \tag{6}$$

$$\underline{\tilde{v}}_{N+1}=-\tilde{\nu}\boldsymbol{C}_N^{-1}\underline{v}_{N+1} \tag{7}$$

$$\tilde{\boldsymbol{C}}=\boldsymbol{C}_N^{-1}+\frac{1}{\tilde{\nu}}\underline{\tilde{v}}_N+\underline{\tilde{v}}_{N+1}^T. \tag{8}$$

By substitution of $\boldsymbol{C}_{N+1}^{-1}$ into (5), the distribution of $y(N+1)$ given the input vector $\psi(N+1)$ can be written [14], [15]

$$P(y(N+1)|D,C(\cdot),\psi(N+1))=\frac{1}{Q}e^{-\frac{1}{2}\frac{(y(N+1)-\mu_{\hat{y}(N+1)})^2}{\sigma^2_{\hat{y}(N+1)}}} \tag{9}$$

where

$$\mu_{\hat{y}(N+1)}=\underline{v}_{N+1}^T\boldsymbol{C}_N^{-1}\underline{y}_N \tag{10}$$

is the mean predicted output for the new input vector and

$$\sigma^2_{\hat{y}(N+1)}=\nu-\underline{v}_{N+1}^T\boldsymbol{C}_N^{-1}\underline{v}_{N+1} \tag{11}$$

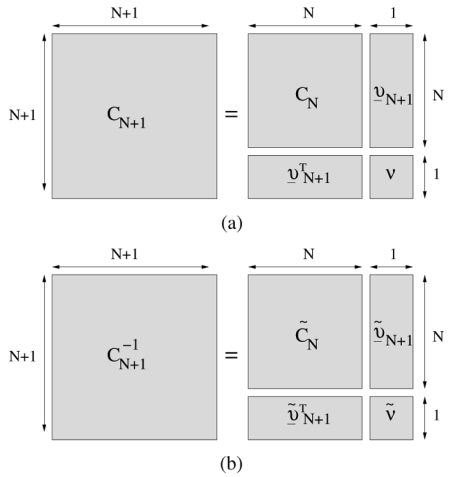

Fig. 2. Construction of the covariance matrix $\boldsymbol{C}_{N+1}$.

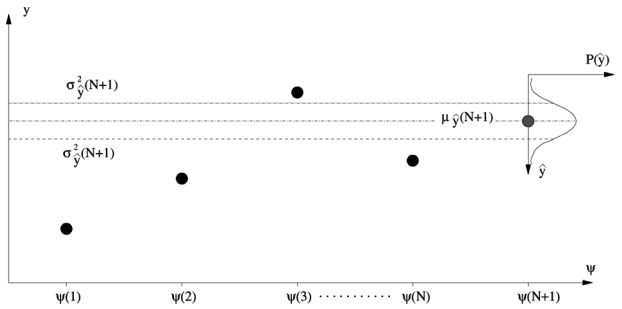

Fig. 3. Predicted output $\hat{y}(N+1)$ and its Gaussian distribution for the input $\psi(N+1)$ and the training set $\psi(1),\ldots,\psi(N)$. The uncertainty of the predicted output is defined by the variance of the Gaussian distribution (dotted lines).

is the variance of the prediction. Therefore, $\mu_{\hat{y}(N+1)}$ is the mean predicted output for the input vector $\hat{y}_{N+1}$ and $\sigma^2_{\hat{y}(N+1)}$ is the variance of this prediction. Put in a more meaningful way, given a new input vector $\psi(N+1)$, the predicted model output $\hat{y}_{N+1}$ is the mean of the Gaussian distribution: $\hat{y} = \mu_{\hat{y}(N+1)}$. The uncertainty of this prediction is given by the variance of the Gaussian distribution: $\sigma^2_{\hat{y}(N+1)}$.

An example given a training set $\psi(1),\psi(2),\ldots,\psi(N)$ with a Gaussian distribution is outlined in Fig. 3. Given the training set and the GP model, a predicted output for the input $\psi(N+1)$ and the certainty of that prediction defined by its variance are calculated from (10) and (11). This prediction and its certainty are represented in this figure by the Gaussian curve at $\psi(N+1)$. The width of the Gaussian curve would be narrower or wider if $\psi(N+1)$ was easier or more difficult to predict.

B. Covariance Function

The GP model depends entirely on the covariance matrix $\boldsymbol{C}$. This matrix, in turn, is produced from the model inputs by the covariance function $C(\cdot)$. One covariance function which has proven to give reliable results and has been widely used in practice is given as

$$C(\psi(m),\psi(n)) = \theta_0 e^{-\frac{1}{2}\sum_{l=1}^{d}\theta_l(\psi_l(m)-\psi_l(n))^2} + \theta_\eta \delta(m,n) \quad (12)$$

where $\underline{\theta} = [\theta_0,\theta_1,\ldots,\theta_d,\theta_\eta]^T$ is the vector of hyperparameters, d is the dimension of the input space and $\delta(m,n)$ is the Kronecker delta function defined as

$$\delta(m,n) = \begin{cases} 1, \text{for } m = n \\ 0, \text{ for } m \neq n. \end{cases} \quad (13)$$

The hyperparameters $\theta_1,\ldots,\theta_d$ correspond to distance measures for each of the d input dimensions. The θ_η hyperparameter is the estimate of the noise variance. Hyperparameter θ_0 controls the overall scale of the local correlation.

C. Training the Gaussian Model

The hyperparameters can be provided as *a priori* information in some cases. In the cases where this is not possible, the result to the following integral must be found in order to obtain the model hyperparameters

$$P(y(N+1)|\underline{\psi}(N+1),D,C(\cdot)) = \int P(y(N+1)|\underline{\psi}(N+1),D,C(\cdot),\underline{\theta})P(\underline{\theta}|C(\cdot))d\underline{\theta}. \quad (14)$$

In most cases, this integral is analytically unsolvable and for its solution two methods have been proposed; the Monte Carlo method and a maximum likelihood approach. While the Monte Carlo method can produce better results for small data sets, it requires large memory storage and a long computation time. For a training set of 64 samples, very close to the 66 sample training set used in this application, Rasmussen quoted 0.25 and 32 min of CPU time for the maximum likelihood and the Monte Carlo approaches, respectively, for a training and prediction routine [16]. Though processing power has increased since Rasmussen's tests in 1996, computation time is still much longer for the Monte Carlo approach, and given that their modeling performance is similar, the maximum likelihood approach is the obvious choice for the estimation of (14) for a real-time application such as neonatal seizure detection. Furthermore, the maximum likelihood approach uses matrix manipulation to minimize the likelihood function. This process is much more efficient for use with reduced instruction set computer (RISC) processors which could be used in the real-time implementation of the algorithm. More information on this maximum likelihood approach can be found in [11], [12], and [16].

D. Advantages of Gaussian Modeling

The level of confidence of parametric models is most frequently estimated based on the uncertainty of the parameters of the model and does not take into account model structure or the distance of the current input from the training data [17], [18], [19]. This essentially means that the model is incapable of detecting if its prediction is valid for a given input. The uncertainty of the Gaussian model carries more information. In the Gaussian model, the prediction is accompanied by a variance measure, shown in (14), based on the model and the distance of the input space from the training set of data (measured in the input space). The larger the distance between the current input point and the training data, the greater the model uncertainty of the prediction. This is important when a nonlinear function is modeled based

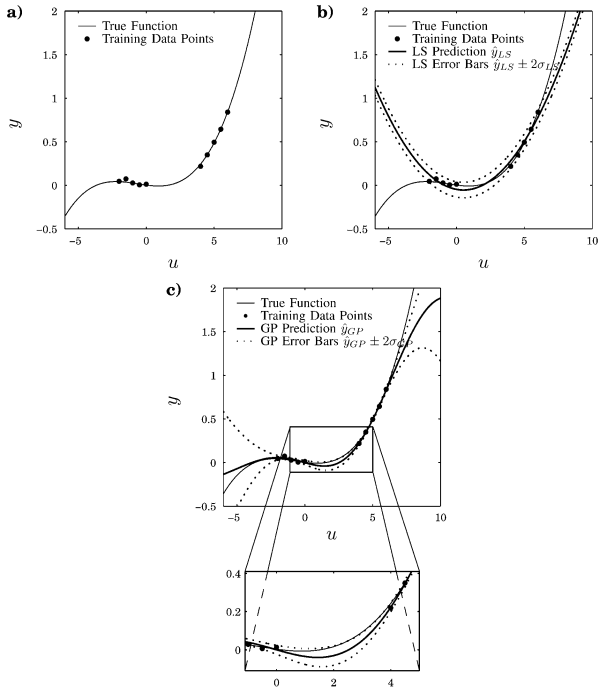

Fig. 4. (a) Training set of data generated by the nonlinear function. (b) Prediction of the parametric model $\hat{y}_{LS} = 0.0277u^2 - 0.0285u - 0.0455$ and its error bars. The error bars do not depend on the distance of the input point from the training data. (c) GP prediction and its error bars. The error bars gets wider away from the training data. Note the lack of data in the region $0 \leq u \leq 4$. The prediction there is less accurate, which is indicated with wider error bars.

on locally collected data (see below). The uncertainty of prediction also depends, of course, on the order of the model with higher order models generally producing predictions with lower uncertainty (problems with higher order models overfitting data aside).

An example of the difference between the uncertainty of a parametric least squared regression approach and a GP approach is shown in Fig. 4. The first plot shows a static nonlinear function $y = f(u) + \epsilon$, where ϵ is a white, zero-mean Gaussian noise, and ten points from the function are chosen to train a linear regression model and a GP model.

A second-order polynomial is fitted to the data using a least squares approach [20] and the resulting function estimate $\hat{y}_{LS}$ is shown in Fig. 4(b). The error bars $\hat{y}_{LS} \pm 2\sigma_{LS}$ are also shown, where σ^2_{LS} are given as [20]

$$\sigma^2_{LS} = \frac{1}{N}\sum_{k=1}^{N}(e(k) - \bar{e})^2 \tag{15}$$

where $e(k)$ is an error of the kth prediction point and $\bar{E}$ is an average of the prediction error over the N prediction points. While the error between the true function and the least squares approximation increases away from the training data, the certainty provided by the model parameters remains unchanged over the entire range, providing only a constant measure of uncertainty in the model's prediction, regardless of the distance of the input from the training data.

A GP model is also trained using the same training data and a similar plot is produced in Fig. 4(c). The variance of the GP prediction depends not only on the parameters of the model, but also on the distance of the input from the training data. The error bars in this case are given by $\hat{y}_{\mathrm{GP}} \pm 2\sigma_{\mathrm{GP}}$, with σ^2_{GP} calculated by (14). Note that, when the input point is far from the training data, the error bars widen, showing an increase in the uncertainty of the prediction. So, while the approximated functions separate from the true function outside the training data with both the least squares and Gaussian model approaches, only the Gaussian model provides an additional measure of uncertainty in the prediction. This additional information is a great advantage to using GP models.

E. Limitations of GP Models

While there are obvious advantages to its use, there are inherent limitations to GP modeling. In some instances, such as automatic control design [11], a level of interpretability of the underlying system from information gained from the model is desirable. As with neural networks, GP models are *black box* in nature and, therefore, lack this transparency.

As described in Section III-A, making a prediction with a GP model requires the inversion of an $N \times N$ matrix, which has a computational cost of $\mathcal{O}(N^3)$. This computation is also required for training. Hence, as the size of the data sets increase, the computational load becomes very large.

IV. Seizure Detection Using GP Modeling

The most straightforward model based seizure detection technique is to derive a model from a section of typical nonseizure EEG and to use this to determine when the EEG activity diverges from the model behaviour. However, it has been discussed in Section II that the neonatal EEG is constantly changing in character, and so it is clear that to choose a typical nonseizure EEG segment, or indeed a typical seizure segment, to train the model would be an unreasonable task. Furthermore, for a reliable, real-time neonatal seizure detection system, the measure used to indicate seizure events must be suitable for use across a wide range of patients. These patients may display very different EEG characteristics, meaning that selecting a standard EEG model is not a viable option.

Hence, in this paper, it is proposed to use the characteristics of a GP model, constantly retrained on new EEG data, to indicate seizure events. There are two ways in which information about the EEG signal can be extracted from the GP model; the variance (or certainty) of the predicted output of the model and the information contained in the hyperparameters of the model. Details of these two approaches are given in the following sections.

A. Variance Approach

As discussed previously, the GP model produces a predicted output along with a measure of the certainty of this prediction for a particular input point. This variance value depends on the distance of the input point from the training data and how well the model has been able to fit the training data. For example, if the model is trained on a random signal, the certainty of a predicted output will be low and, hence, the variance high. Alternatively, if the model is trained on a deterministic signal, the certainty of the predicted output will be high, and the variance will

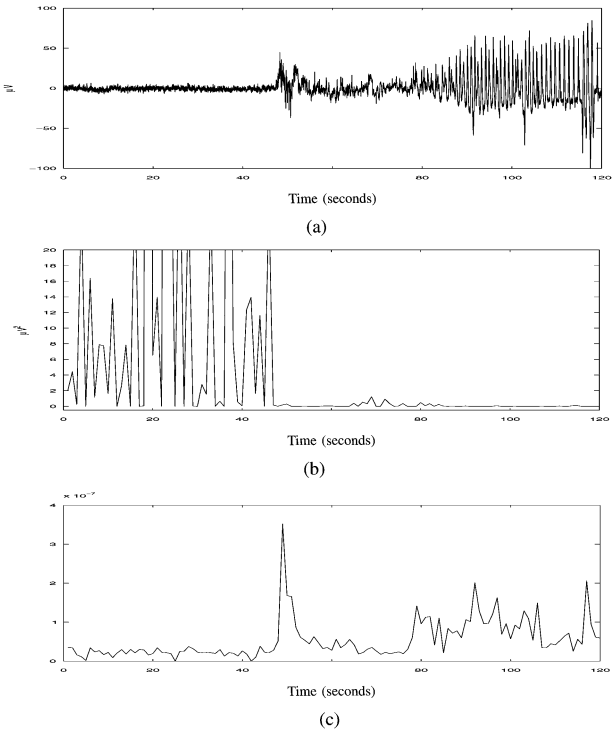

Fig. 5. (a) Section of neonatal EEG at the onset of a seizure. (b) Prediction variance. (c) Hyperparameter ratio.

be low. In Section II, it was shown that the neonatal seizure EEG is more repetitive and deterministic than the nonseizure EEG. Therefore, the variance of the predicted model output could be used to indicate changes in structure of the neonatal EEG signal; a change from nonseizure to seizure should be accompanied by a drop in the prediction variance.

The data for training the GP model is obtained from the EEG using a sliding window of length 1 s (data sampling frequency= 66 Hz; therefore, $N = 66$) with an overlap of 1/6 s. The short window length is used as computation time increases greatly with an increase in the number of training points, an important consideration in the design of a real-time seizure detection system, and the GP model is known to obtain relatively good results with a small data set [11], [12]. The input matrix is generated using Taken's method of delays [21] with a time delay of one sample and the embedding dimension set by the order of the model, d. All but the last data vector are used for training the GP model. The last vector is then used to calculate a one step ahead prediction from the trained model. The variance of this prediction represents the amount of determinism in the EEG signal.

In Fig. 5(a), a section of neonatal EEG of length 2 min is shown. A seizure begins at $t \approx 50$ s. The data is analyzed using the approach outlined above. Fig. 5(b) shows the associated prediction variance for the same EEG segment. For the nonseizure EEG the variance is correspondingly high, showing its lack of determinism. As the EEG enters seizure the variance drops significantly and remains low for the duration of the seizure. This change in variance is an indicator of the more deterministic nature of the seizure EEG. The opposite effect occurs at the cessation of a seizure.

B. Hyperparameter Approach

In Section III, the hyperparameters $\underline{\theta} = [\theta_0, \theta_1, \ldots, \theta_d, \theta_\eta]^T$ were introduced. θ_0 relates to the local correlation of the input matrix and θ_η relates to the noise in the data. When modeling neonatal EEG data, the model hyperparameters change from one EEG segment to the next. At a point when the EEG characteristics change dramatically, such as at seizure onset, the θ_0 hyperparameter reflects this change as the level of determinism in the signal changes. This hyperparameter reflects most of the change in the EEG while the rest of the information is spread across the other hyperparameters. Meanwhile, as the level of noise in the signal remains relatively steady over this change, the θ_η hyperparameter remains consistent and so is chosen as a good reference point for θ_0. Therefore, the ratio of the magnitudes of the hyperparameters $|\theta_0/\theta_\eta|$ is an indicator of the level of determinism in the signal. This will be referred to as the *hyperparameter ratio*.

Another reason to choose the θ_0 hyperparameter is that even with a badly chosen model order (model order is discussed in the next section), θ_0 will still show distinct changes as the characteristics of the signal change. This method, therefore, has an advantage over other modeling methods in that the hyperparameter ratio is robust to changes in model-order selection which may degrade the performance of other modeling techniques considerably.

This measure should prove useful as an indicator of a change in determinism of the EEG, which accompanies seizure events. The same moving window and input matrix generation approach as that for the prediction variance approach is used to obtain training data for the GP model. In Fig. 5(c), the effect of the transition from nonseizure to seizure EEG is shown for the $|\theta_0/\theta_\eta|$ ratio. $|\theta_0/\theta_\eta|$ rises as expected as the EEG enters seizure, corresponding to the increase in the amount of determinism in the EEG.

V. Experimental Setup

All EEG data was collected from newborn babies with seizures in the neonatal intensive care units of the Amalgamated Maternity Hospitals, Cork, Ireland. Written consent was obtained from the parents of each baby studied and this paper had full ethical approval from the Ethics Committees of the hospitals. A Viasys NicOne video-EEG system was used to record 12 channels of EEG using the 10–20 system of electrode placement modified for neonates ($F_4 - C_4, C_4 - P_4, P_4 - O_2$, $F_3 - C_3, C_3 - P_3, P_3 - O_1$, $T_4 - C_4, C_4 - C_z, C_z - C_3, C_3 - T_3, T_4 - O_2, T_3 - O_1$). Two additional channels were used to record ECG and respiration. A video recording was also made of each neonate for the duration of the paper. All digitized EEG data was then converted to EDF file format [22]. A clinical neurophysiologist identified and classified all periods of seizure activity in each EEG recording.

The data used for testing the GP modeling approaches consisted of data from four patients, one female, three male, all aged between 4 and 72 h, all full-term and with hypoxic ischaemic encephalopathy (HIE). The details of each recording are shown in Table I. The test EEG totaled 51 h, containing 277 seizures. The data was initially recorded at 200 Hz, but was decimated to

TABLE I
CHARACTERISTICS OF THE RECORDINGS IN THE TEST EEG SET. ALL PATIENTS WERE FULL-TERM (GESTATIONAL AGE OF >40 WEEKS. *REC* IS THE RECORD ID, *LENGTH* IS THE RECORD LENGTH IN HOURS, *CHAN* IS THE NUMBER OF CHANNELS, *SEIZ* IS THE NUMBER OF SEIZURES, AND *DUR* IS THE MEAN DURATION OF SEIZURE IN MINUTES

Rec	Sex	Outcome	Length	Chan	Seiz	Dur
A	F	Deceased	10	8	71	1.71
B	M	Disabilities	24	9	156	5.27
C	M	Disabilities	12	9	29	2.15
D	M	Normal	5	6	23	1.02

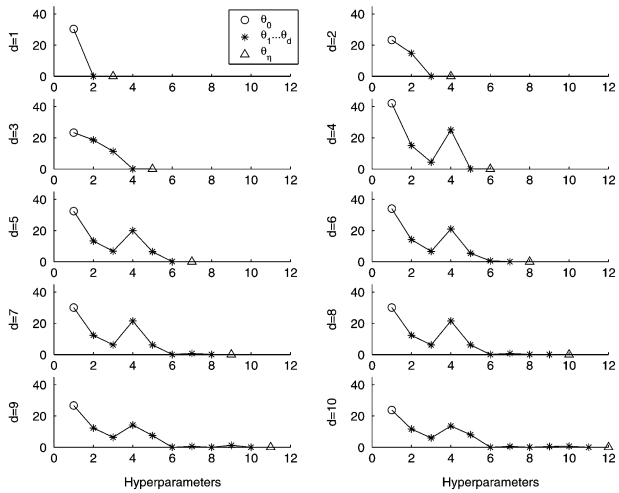

Fig. 6. Hyperparameters for increasing GP model order for a nonseizure neonatal EEG segment.

66 Hz before analysis to speed up the modeling process. This is a reasonable procedure given that the frequencies of interest for neonatal seizure detection range from approximately 0.1 to 30 Hz, giving a Nyquist frequency of $\approx$60 Hz. The data was normalized to have zero mean and lie between -1 and 1 in line with the assumptions made in Section III. As mentioned previously, the EEG data was analyzed with a sliding window of length 1 s and an overlap of 1/6 s. The window overlap ensures that no information is lost due to windowing and also performs a smoothing of the output. All 12 channels were analyzed in this paper.

For each window of EEG, a Gaussian model was trained. To determine the optimum model order for neonatal EEG seizure detection, hyperparameters for the GP models were estimated for neonatal EEG for increasing model dimension. As the model order increases, a point is reached where the hyperparameters remain stable and the higher hyperparameters are negligible. Models were estimated for 1 h of EEG, including nonseizure and seizure EEG from four neonates. The hyperparameters were examined for increasing model order and a sixth-order model was determined to be adequate to model the EEG while keeping computational cost as low as possible. An example of the hyperparameters of increasing order models is shown in Fig. 6. This particular example shows consistent hyperparameters from an order (*dim*) of 5 and up.

To confirm the choice of model order, a method of determining the dimension of a system, which is the same as the order of the model used to represent that system, a nonlinear dynamics system theory approach known as the *false nearest neighbors* approach [23], [24] is implemented. The EEG data is embedded by Taken's method of delays for increasing dimensions. The percentage of neighboring points which are neighbors only by virtue of the projection into too low a dimension is calculated. As the embedding dimension increases, the number of these points drops until a noise floor is reached. For the case of the test EEG, the percentage of false neighbors for each dimension is calculated and a reduction to below the 5% mark is achieved for all the test EEG segments by a dimension of 6, as shown in Fig. 7, confirming the model-order choice made from the hyerparameter test described above. Similar model orders were used in other work on epileptic seizures [25].

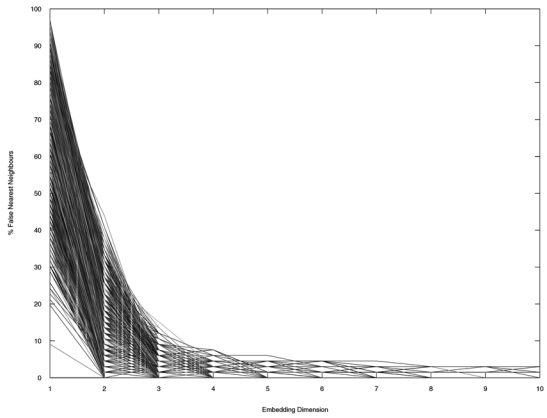

Fig. 7. Percentage of points in the embedded space which are false neighbors for increasing embedding dimension. Each line here represents the result for one segment from the test EEG.

From this model, the variance of a one-step ahead prediction and the hyperparameter ratio θ_0/θ_η were calculated. While it would be possible to build the input matrix using spacial embedding (the input matrix being made up of the EEG from each channel) rather than time delay embedding, there are a number of reasons why it was chosen to calculate the measures on a single electrode at a time. First, while there are guidelines as to the number of electrodes to use for neonatal EEG monitoring [26], often the number of electrodes differs from recording to recording. The recording from an electrode may also be void due to poor contact with the patient or artifact. Hence, the structure of the input matrix could be very different from patient to patient making development of an expert system very difficult. Second, the single-channel analysis is used in the system proposed in [7], in which multichannel analysis has already been incorporated by mean of independent component analysis (ICA). Also, most of the other measures discussed in this work are designed for single-channel analysis, and so it would be an unfair comparison of performance if it were applied to the Gaussian model. Finally, neonatal seizures may be focal in nature, only appearing at a single or a low number of electrodes. This would essentially mean that the input matrix would be partly made up of nonseizure EEG and partly of seizure EEG. This would lead to difficulties in defining results and in structuring classification networks for the system.

TABLE II
MEAN AND STANDARD DEVIATION VALUES FOR THE PREDICTION VARIANCE AND HYPERPARAMETER RATIO FOR SEIZURE AND NONSEIZURE EEG

Predict Var	Mean	Std
Seiz	0.7613	18.259
Non-Seiz	15.3678	79.8151

Hyper Ratio	Mean	Std
Seiz	0.3014	0.2285
Non-Seiz	0.3895	0.2467

VI. RESULTS

As expected, during seizure events, the prediction variance decreased by several orders of magnitude, highlighting the more repetitive, organized nature of the EEG during seizure. Correspondingly, the GP model shows an increase in the hyperparameter ratio as the EEG changes from nonseizure to seizure, though it is a much smaller change. The mean and standard deviation of each measure for nonseizure and seizure EEG, for all 51 h of test EEG, is given in Table II.

To give some statistical significance to the results presented in Table II, an analysis of the distributions of the nonseizure and seizure values is carried out to test whether the values are statistically different. Obviously, if the values are not statistically different, they will be very poor indicators of neonatal seizures. The classical test for determining the difference of two data sets is the *t-test*. However, there are two assumptions to the t-test which are violated here. First, from Table II, the variances of the data sets are not equal (variance is the square of the standard deviation). Second, the data is not of a normal distribution. Because of these violations, a nonparametric analysis must be made and the corresponding nonparametric test, the *Wilcoxon rank-sum* test [27], is used. The results show that both the prediction variance and hyperparameter ratio have statistically different distributions for nonseizure and seizure EEG at the 5% level of significance. In fact, pushed beyong the standard 5% level, the values are significantly different down to the 1% level of significance (less than a 1% chance that they came from the same distribution). To give some means of comparing their performances, the rank-sum statistic, upon which the Wilcoxon rank-sum test is based, is quoted; the higher the rank-sum statistic, the greater the difference in distributions. The rank-sum statistic for the prediction variance approach is $7.1211e^9$ and for the hyperparameter approach is $6.7263e^9$. The lower value for the hyperparameter approach is to be expected given the values in Table II. Boxplots of the distributions of the two measures are given in Fig. 8. For the prediction variance approach distribution, shown in Fig. 8(a), the seizure prediction variance values are all extremely low, rarely rising above $20\mu V^2$. The nonseizure values are considerably higher, rising up to over $100\mu V^2$. This considerable difference in nonseizure and seizure values should allow for accurate classification of neonatal EEG. Fig. 8(b) shows that the difference in the distributions of the hyperparameter approach for nonseizure and seizure EEG is not as large. The nonseizure ratio values are mostly spread over a larger range than the seizure values, but

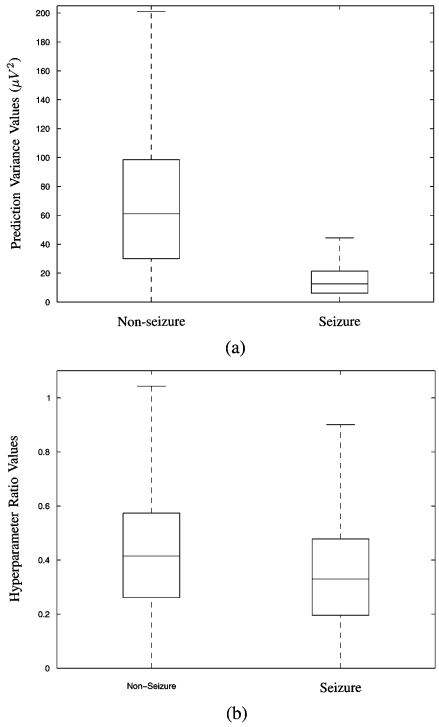

Fig. 8. Distributions of the GP modeling measures. The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. (a) Distribution of the prediction variance (in μV).

TABLE III
MEAN AND STANDARD DEVIATION VALUES FOR THE SIGNAL VARIANCE FOR SEIZURE AND NONSEIZURE EEG

Signal Variance	Mean	Std
Seiz	$1.3123e^4$	$6.1439e^4$
Non-Seiz	$1.2209e^4$	$6.4344e^4$

their overlapping distributions mean that while some classification success is possible, use of this measure alone would not appear to be sufficient for accurate neonatal seizure detection.

As a further reference for the performance of the GP methods, the signal variance (not to be confused with the prediction variance of the GP model) is also calculated. Viewing the raw EEG in Figs. 1 and 5, it might appear as though the increase in the variance of the signal would be enough to detect the presence of seizures. If this were the case, there would be no need for the complicated process of GP modeling and EEG analysis could be carried out in a much smaller time frame. To determine if the variance can indeed be used to distinguish nonseizure from seizure neonatal EEG, the same statistical tests completed above for the GP approaches have been carried out with the signal variance. The mean and standard deviation values for nonseizure and seizure neonatal EEG are given in Table III. The Wilcoxon rank-sum test gives a rank-sum statistic of $1.2944e^8$, about 50 times smaller than that for the GP approaches. Viewing the distribution of the seizure and nonseizure values in Fig. 9, no considerable difference exists between the nonseizure and seizure results. Therefore, the reliable detection of neonatal seizures requires more advanced measures such as the GP modeling approaches proposed here.

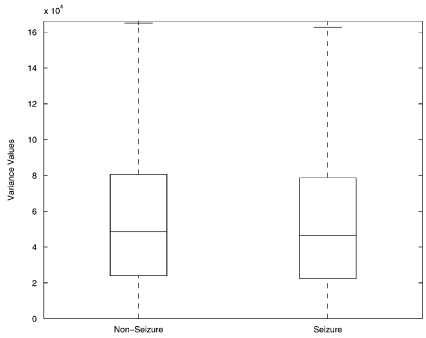

Fig. 9. Distribution of the variance of the test nonseizure and seizure neonatal EEG.

VII. Comparison With Other EEG Measures

In past studies, it was noted that for reliable neonatal seizure detection, an EEG feature set encompassing measures from more than one area of signal processing would be required [6]. If the GP model methods that are introduced here are to be of use in a seizure detection system, they must prove to be good indicators of seizure and must provide a different insight into the EEG behaviour than features that have been already developed.

In this section, a set of ten features from various areas of signal processing is tested to provide a benchmark for the GP modeling approaches. A comparison is performed with a different modeling approach, measures that have recently been proposed for EEG analysis, and measures already in clinical use today. Firstly, an autoregressive modeling approach is outlined. Then the other measures are briefly discussed. Finally, the performance of the GP measures are compared to those previously proposed measures using neural network classification.

A. Autoregressive Modeling

Autoregressive (AR) modeling is a widely used technique in signal processing and has previously been used for EEG analysis [28]. In the normal course of system identification, once a model has been developed, its effectiveness can be tested by validating the model on some data which was not used to derive the model. In this section, this method of validation is utilized to determine the level of determinism in the EEG signal and, hence, to indicate seizures. The performance of this method is compared to that of the GP model approaches.

The following dth-order autoregressive (AR) model is assumed in this paper

$$y(k) = \sum_{i=1}^{d} a_i y(k-i) + \eta(k). \tag{16}$$

Here, y is the output of the model, d is the model order, a_i are the model parameters and η is Gaussian white noise. The model parameters are obtained by minimizing the sum of least-squares error for the forward model and for a time-reversed model. This approach is known as the modified covariance method. It may be of note that d is the order of the AR model, whereas it is the dimension of the input space of the GP model. In time series modeling this equates to the same measure, and so the letter d is used in this paper for both.

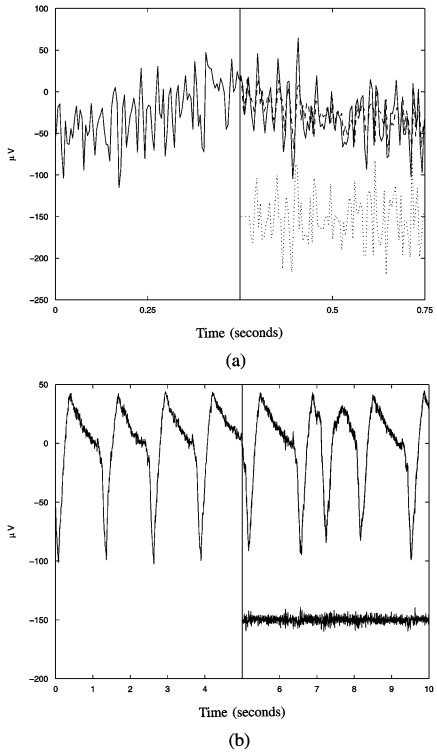

Fig. 10. AR fit examples for nonseizure and seizure EEG. In this example, the first 5 s is used for training and the second 5 s is used for validation (1 s is used in later tests). The measured data is shown by the solid line, the simulated output is shown by the dashed line and the error is shown by the dotted line. The error is shifted by $-150\,\mu$V to aid visualization. The time resolution on the nonseizure plot is also increased for visual purposes. For the nonseizure EEG, the fit was calculated at $\approx 32\%$. For the seizure EEG, the fit is so good as to not be seperable in the figure and the error is noticably reduced. The fit for this seizure segment was calculated at $\approx 91\%$. (a) AR model trained on a section of nonseizure EEG. (b) AR model trained on a section of seizure EEG.

Once the parameters of an AR model have been calculated, its ability to fit another data set can be analyzed for validation of the model. If the two data sets are not similar, the fit will be poor. Therefore, as shown in Fig. 10(a), an AR model trained on a section of nonseizure EEG provides a poor fit when validated on another section of nonseizure EEG. However, seizure EEG is more organized and repetitive in nature, and, therefore, as highlighted in Fig. 10(b), a model trained using one section of seizure EEG provides a good fit on another section of seizure EEG. It is clear from these figures that the error for the seizure EEG is much smaller than for the nonseizure EEG.

The percentage of the output variation that is explained by the model is given by

$$\text{fit} = 100 \times \left(1 - \frac{\lambda(\hat{y} - y)}{\lambda(y - \bar{y})}\right)\% \tag{17}$$

where y is the validation data, $\hat{y}$ is the one step ahead predicted output, $\bar{y}$ is the mean of the validation data and $\lambda(x) = \sqrt{\sum x^2}$ is the *norm* of x.

Using this approach, the EEG is windowed in a similar manner to that described in Section V, though with the first half of each window used for training and the second half of the window for validation. For a direct comparison, the window length is made so that the AR model has the same length

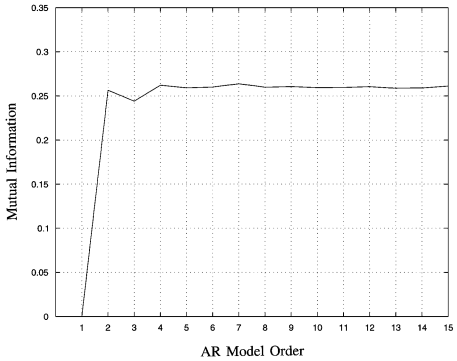

Fig. 11. Mutual information of the fit of various order AR models with seizures in test EEG. A higher mutual information.

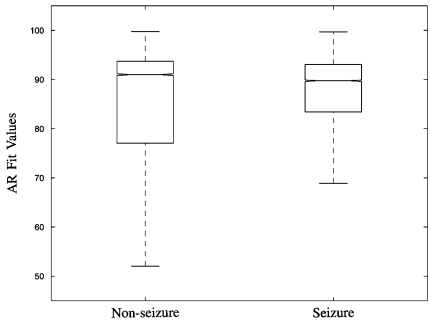

Fig. 12. Distribution of the AR modeling feature for nonseizure and seizure EEG.

training set as the GP approaches (so both models are trained on 1 s of EEG). Simulations were carried out for a range of AR model orders from two to 15, and the mutual information between the seizure events and the AR model fit for varying model order was found (Fig. 11). The greater the change in the AR model fit with seizure events, the greater the mutual information. The performance is similar for orders $d = 4$ and above (Fig. 11). A seventh-order model gave the largest mutual information with the seizure times and was very close to the choice of GP model order (see Section V) and is, hence, used here for comparison. A performance comparison is carried out based on the ANOVA test introduced in Section VI. The ANOVA test shows that the AR model approach also produces values for seizure and nonseizure EEG that are statistically different at the 5% level of confidence. The distribution of the values for nonseizure and seizure EEG, shown in Fig. 12, shows a similar difference in nonseizure and seizure values as those for the hyperparameter ratio, given in Fig. 8(b). Therefore, similar performance is expected from the AR approach as to the GP hyperparameter approach.

B. Other EEG Measures

Initial studies on EEG analysis depended on frequency information to highlight seizure activity [29], [30]. While it became clear in further studies that these measures alone were not enough to reliably detect neonatal seizures, it was also clear that there were obvious changes in the frequency spectrum during seizure events [6]. For this paper, the intensity weighted mean frequency and bandwidth measures are calculated for comparison [6], [31]. These frequency-calculation measures provide a better estimation of the dominant frequency and bandwidth of a signal than the more simple maximum peak detection methods [6].

Measures of the spatial and dynamic properties of the attractor of sections of the EEG signal, known as nonlinear dynamic systems theory or chaos theory, have been investigated previously and are discussed in [32]. A chaos theory measure known as the KY dimension which measures the spatial spread of the attractor and the Shannon Entropy with which it is compared in [32] are included in this feature set. Takens method of delays is used to embed the data for analysis with an embedding dimension of six (see Fig. 7) and a time delay of one sample. These parameter values ensure that enough data is taken to determine the signal characteristics and that no information is lost by using a larger time delay.

Wavelet analysis is also a much researched area in recent EEG studies. These techniques allow the tracking of changes in the frequency content of the signal over time and are free from some of the assumptions of the more traditional Fourier transform method of frequency analysis. The wavelet analysis seizure detection measure described in [33], in which the relative energies in the wavelet sub-bands are compared to detect changes in frequency content, is also included in this feature set.

Singular value decomposition (SVD) is used in signal processing to determine the number of major components of a signal and is particularly suited to the analysis of quasi-periodic signals in noise [34], which is the nature of the EEG classification problem. An approach to quantifying the amount of power in the first of these singular values, known as the singular value fraction (SVF) is given in [35]. This analysis tool is included in the feature test set. Again, before analysis by this measure, the data must be embedded and Takens method of delays is used as above.

Equipment currently in clinical use for the analysis of EEG for estimating the depth of anaesthesia employs the measure known as spectral entropy. This measure quantifies the amount of information in the signal which can be determined from its frequency content [36]. As this measure is in clinical use as an EEG feature, it was included in the feature set to compare its performance with those measures proposed in this paper.

C. Neural Network Classification

To assess the effectiveness of the GP measures in indicating the presence of neonatal seizures, a neural network is trained with each measure, from nonseizure and seizure data from the EEG test data, and used to classify test EEG from the same patient. The values from each measure are normalized to fall between -1 and 1 before being fed into the neural network and those features calculated over 10 s segments are interpolated to a resolution of 1 s so that each test has the same number of inputs. A two-layer neural network is used, using a hyperbolic tangent sigmoid transfer function in the first layer and a pure linear transfer function in the second layer, with 35 neurons and one neuron respectively chosen through trial and error tests. This network was chosen after tests were carried out over a wide range of network sizes and characteristics, and this configuration produced the optimum results. The complexity of the classifier needed to produce high performance shows one

TABLE IV
CLASSIFICATION ACCURACY OF EACH EEG MEASURE

EEG Measure	Accuracy (%)
Gaussian Variance	82.8
Wavelet Energy	79.1
Intensity-weighted Bandwidth	78.6
AR Model Fit	78.0
Entropy	77.2
Singular Value Fraction	76.5
Gaussian Hyperparameter Ratio	62.7
Spectral Entropy	62.4
Intensity-weighted Mean Frequency	62.1
Kaplan-Yorke Dimension	52.7

of the main problems in EEG classification: EEG data is very much ill-posed. This also goes to show why straightforward classification routines previously attempted fail to provide accurate seizure detection [6]. Work on optimum classifiers for neonatal seizure detection is currently ongoing, with Bayesian and linear and quadratic discriminant classifiers [37] currently being tested. Results are based on the standard sensitivity (the number of seizure segments classified correctly) and specificity (the number of nonseizure epochs classified correctly) measures. A sweep of threshold values is carried out at the output to determine the optimal results. As the output threshold increases, the sensitivity will decrease and the specificity increase. The optimal result is deemed to be where the two cross. This gives the highest combination of specificity and sensitivity (referred to as the classification accuracy below).

While these measures and a neural network classifier work together to make up the essentials of a neonatal seizure detection system, it should be made clear that this work is only a section of a larger seizure detection system proposed in [7], and that the results presented here do not indicate the performance of the entire system. These results are only presented as a means of comparison of the different measures available for neonatal EEG analysis.

D. Results

All ten features were calculated for 51 h of test neonatal EEG. The EEG data was windowed similarly to the approach taken with the Gaussian and AR measures but with a longer window of 10 s and an overlap of 2.5 s. The longer window is needed by most measures to obtain an adequate representation of the signal. The sensitivity/specificity at which the optimal performance was achieved by the neural network for each of the EEG measures is given in Table IV.

These results show the power of the prediction variance approach in indicating the presence of neonatal seizures. This measure alone achieves a classification accuracy of 82.8% for the data set used for this paper, a full 3.7% ahead of the rest of the features. As expected from the rank-sum tests in Section VI, the Gaussian hyperparameter approach achieves a much poorer performance, achieving a classification accuracy of only 62.7%. The wavelet energy, intensity-weighted bandwidth, AR model fit, entropy and SVF measures also performed well ranging from 79.1% to 76.5%.

These results show that the GP model prediction variance approach has a considerable performance advantage over the AR approach and other previously proposed neonatal seizure detection measures. The spectral entropy measure, which is in current clinical use as an EEG measure of anaesthetic depth, performed poorly in these tests as an indicator of neonatal seizure, achieving a classification accuracy of only 62.4%. This shows the high performance of the GP prediction variance approach in comparison to tried and trusted EEG analysis techniques and proves that the GP approaches show good promise for use in a clinical environment.

The hyperparameter ratio method proved less successful only achieving a classification accuracy of 62.7%. It is believed that the performance difference is due to the fact that while the θ_0 hyperparameter shows obvious changes during seizure events, there is some information which is only modeled by the other hyperparameters. This information loss may be relatively small but is enough to degrade the performance of the hyperparameter ratio considerably in neonatal EEG classification. The hyperparameter ratio approach is, however, robust against inaccurate selection of the model dimension d. This advantage makes it useful in applications of real-time signal analysis where the dimension of the underlying system may not be consistent.

A disadvantage to using the Guassian process modeling methods is that computation times are high. The GP approach takes 1.6172 s to calculate one hyperparameter ratio and prediction variance for a second of EEG data at 66 Hz (hyperparameters must be calculated before prediction can be found). In comparison, the AR fit approach takes only 0.0329 s for the same amount of data (times taken for *Matlab* implementation on a 3.2 GHz *Intel Pentuim* 4 CPU with 1-Gb RAM). This is the biggest drawback to the use of this method for real-time seizure detection. However, a C or hardware implementation would decrease the computation time considerably.

There are other areas where future research may allow for higher detection rates. A solid theoretical foundation for the selection of input space dimension d should be introduced so that a more accurate model can be estimated while keeping computation times low. Also, a standard covariance function $C(\cdot)$ is used in this analysis where a different covariance function would produce a different set of hyperparameters which would have a different interpretation. Therefore, the choice of a function which is specifically suited to the seizure detection task may produce more accurate results for the hyperparameter ratio measure. This could also reduce computational costs further.

While, in this analysis, it was chosen to analyze the EEG with the Gaussian modeling approaches using 1 s segments, it would be interesting to see how the results would be affected using longer segments. Gaussian models can provide robust models from small data sets, but the minimum frequency being analyzed in this case is 1 Hz. While the majority of seizure EEG activity occurs in the 2 to 5 Hz range, seizures can consist of slow wave activity which may lie outside this range. Such analysis would realistically have to wait until a more efficient implementation of the Gaussian modeling techniques is carried out.

As the EEG is nonstationary in nature, parameters such as d and the embedding parameters used for Taken's embedding method change from one segment to the next. Therefore, any

choice of these values is only an educated guess made on the mean value from various tests. In the future, to confirm the results shown here and test the features ability to perform when the parameters used are not optimum, tests with other parameter values of embedding dimensions should also be preformed.

VIII. Conclusion

In this paper, the theory of GP modeling has been applied to the neonatal seizure detection problem. Two measures of EEG structure have been derived from the GP model; the prediction variance and the hyperparameter ratio. During seizure events the prediction variance was seen to reduce dramatically, accompanied by an increase in the ratio of the first hyperparameter to the last. In tests with over 50 h of real neonatal EEG containing seizures, both measures showed a statistical difference in their values from nonseizure to seizure EEG at the 5% significance level. Tests against an AR modeling approach and seven other EEG measures from various fields of signal processing, including one measure which is currently in use as an EEG analysis tool, has shown that in a clinical neonatal seizure detection scenario the GP model measures would be expected to perform well relative to presently available EEG measures. The performance of the prediction variance approach was particularly promising, performing the best out of all the features tested by a significant margin. The hyperparameter ratio approach was less successful, but has other advantages over more traditional EEG modeling methods including its robustness to model-order selection. Moreover, with further processing of the raw EEG signal [7], [38], the ability of the GP model to detect seizures in neonatal EEG should be enhanced even further, while a more efficient implementation and changes to the covariance function used could considerably reduce computation times to provide a quick, accurate means of neonatal seizure detection.

References

[1] M. Lanska, D. Lanska, R. Baumann, and R. Kryscio, "A population-based study of neonatal seizures in Fayette county, Kentucky," *Neurology*, vol. 45, no. 4, pp. 724–732, Apr. 1995.

[2] A. Watkins, W. Szymonowicz, X. Jin, and V. Yu, "Significance of seizures in very low birthweight infants," *Dev. Med. Child Neural.*, vol. 30, no. 2, pp. 162–169, Apr. 1988.

[3] M. S. Scher, *Electroencephalography of the Newborn: Normal and Abnormal Features*, E. Niedermeyer and L. DaSilva, Eds., 4th ed. Baltimore, MD: Williams & Wilkins, 1999, pp. 869–946.

[4] M. McBride, N. Laroia, and R. Guillet, "Electrographic seizures in neonates correlate with poor neurodevelopmental outcome," *Neurology*, vol. 55, pp. 506–513, 2000.

[5] J. Rennie, G. Chorley, G. Boylan, R. Pressler, Y. Nguyen, and R. Hooper, "Non-expert use of the cerebral function monitor for neonatal seizure detection," *Arch. Disease Childhood*, vol. 89, pp. F37–F40, 2004.

[6] S. Paul, G. Boylan, S. Connolly, W. Marnane, and G. Lightbody, "An evaluation of automated neonatal seizure detection methods," *Clin. Neurophysiol.*, vol. 116, no. 7, pp. 1533–1541, Jul. 2005.

[7] S. Paul, G. Boylan, S. Connolly, W. Marnane, and G. Lightbody, "A novel automatic neonatal seizure detection system," in *Proc. IEE Irish Signals and Systems Conf.*, Sep. 2005, pp. 377–382.

[8] S. Patrizi, G. Holmes, M. Orzalesi, and F. Allemand, "Neonatal seizures: Characteristics of EEG ictal activity in preterm and fullterm infants," *Brain Dev.*, vol. 25, no. 6, pp. 427–437, 2003.

[9] R. Bates, M. Sun, M. Scheuer, and R. Sclabassi, "Seizure detection by recurrent backpropagation neural network analysis," in *Proc. IEEE 4th Int. Symp. Uncertainty Modeling and Analysis*, 2003, pp. 312–317.

[10] W. Weng and K. Khorasani, "An adaptive structure neural network with application to EEG automatic seizure detection," *Neural Netw.*, vol. 9, no. 7, pp. 1223–1240, 1996.

[11] G. Gregorcic, "Data-based modeling of nonlinear systems for control," Ph.D. dissertation, Univ. College Cork, Cork, Ireland, 2004.

[12] G. Gregorcic and G. Lightbody, "Gaussian process approaches to nonlinear modelling for control," in *Intelligent Control Systems Using Computational Intelligence Techniques*, A. E. Ruano, Ed. London, U.K.: IEE, 2005, ch. 6, pp. 177–217.

[13] S. Barnett, *Matrix Methods for Engineers and Scientists*. London, U.K.: McGraw-Hill, 1979.

[14] M. Gibbs, "Bayesian Gaussian processes for regression and classification," Ph.D. dissertation, Univ. Cambridge, Cambridge, U.K., 1997.

[15] C. Williams and C. Rasmussen, , M. H. Touretzky and M. Mozer, Eds., "Gaussian processes for regression," in *Advances in Neural Information Processing Systems 8*. Cambridge, MA: MIT Press, 1996, pp. 514–520.

[16] C. Rasmussen, "Evaluation of Gaussian processes and other methods for non-linear regression," Ph.D. dissertation, Univ. Toronto, Toronto, ON, Canada, 1996.

[17] R. Murray-Smith and D. Sbarbaro, "Nonlinear adaptive control using non-parametric gaussian process models," presented at the 15th Int. IFAC Triennial World Congr. Federation of Automatic Control, Barcelona, Spain, Jul. 2002.

[18] R. Murray-Smith, D. Sbarbaro, C. Rasmussen, and A. Girard, "Adaptive, cautious, predictive control with gaussian process priors," presented at the 13th IFAC Symp. System Identification, Rotterdam, The Netherlands, Aug. 2003.

[19] J. Kocijan, A. Girard, B. Banko, and R. Murray-Smith, "Dynamic systems identification with gaussian processes," in *Proc. 4th Mathomod*, Vienna, Austria, 2003, pp. 776–784.

[20] I. Bronshtein, K. Semendyayev, G. Musiol, and H. Muhlig, *Handbook of Mathematics*, 4th ed. New York: Springer, 2004.

[21] F. Takens, *Detecting Strange Attractors in Turbulence*, ser. Lecture Notes Math. Berlin, Germany: Springer, 1981, vol. 898, pp. 366–381.

[22] B. Kemp, European Data Format. [Online]. Available: http://www.edfplus.info

[23] M. B. Kennel, R. Brown, and H. D. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, no. 6, p. 3403, Mar. 1992.

[24] H. D. Abarbanel, "The analysis of observed chaotic data in physical systems," *Rev. Modern Phys.*, vol. 65, no. 4, pp. 1331–1392, Oct. 1993.

[25] L. lasemidis, D. Shiau, J. Sackellares, P. Pardalos, and A. Prasad, "Dynamical resetting of the human brain at epileptic seizures: Application of nonlinear dynamics and global optimization techniques," *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 493–506, 2004.

[26] H. Jasper, "The ten-twenty electrode system of the international federation," *Electroencephalogr. Clin. Neurophysiol.*, vol. 10, pp. 371–375, 1958.

[27] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, no. 6, pp. 80–83, Dec. 1945.

[28] G. Mohammadi, P. Shoushtari, B. M. Ardekani, and M. B. Shamsollahi, "Person identification by using AR model for EEG signals," *Trans. Eng., Comput. Technol.*, vol. 7, pp. 281–285, Feb. 2006.

[29] J. Gotman, D. Flanagan, J. Zhang, and B. Rosenblatt, "Automatic seizure detection in the newborn: Methods and initial evaluation," *Electroenceph. Clin. Neurophysiol.*, vol. 103, pp. 356–362, 1997.

[30] A. Liu, J. Hahn, G. Heldt, and R. Coen, "Detection of neonatal seizures through computerized EEG analysis," *Electroenceph. Clin. Neurophysiol.*, vol. 82, pp. 30–37, 1992.

[31] D. H. Evans and W. N. McDicken, *Doppler Ultrasound. Physics, Instrumentation and Signal Processing*, 2nd ed. New York: Wiley, 2000.

[32] S. Faul, G. Boylan, S. Connolly, W. Marnane, and G. Lightbody, "Chaos theory analysis of the newborn EEG: Is it worth the wait?," in *Proc. IEEE Int. Symp. Intelligent Signal Processing*, Sep. 2005, pp. 381–386.

[33] B. Boashash, M. Mesbah, and P. Colditz, "Newborn EEG seizure pattern characterization using time-frequency analysis," *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, vol. 2, pp. 1041–1044, May 2001.

[34] P. Celka and P. Colditz, "A computer-aided detection of EEG seizures in infants: A singular-spectrum approach and performance comparison," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 5, pp. 452–462, May 2002.

[35] G. Kember and A. Fowler, "A correlation function for choosing time delays in phase portrait reconstructions," *Phys. Lett. A*, vol. 179, pp. 72–80, 1993.

[36] H. Viertio-Oja, V. Maja, M. Sarkelii, P. Talja, N. Tenkanen, H. Tolvanen-Laakso, M. Paloheimo, A. Vakkuri, A. Yli-Hankala, and P. Merilainen, "Description of the Entropy™ algorithm as applied in the Datex-Ohmeda S/5™ entropy module," *Acta Anaethesiol. Scand.*, vol. 48, pp. 154–161, 2004.

[37] B. Greene, R. Reilly, G. Boylan, P. de Chazal, and S. Connolly, "Multichannel EEG based neonatal seizure detection," in *Proc. IEEE 28th Annu. Conf. Engineering in Medicine and Biology Society*, Aug. 2006, pp. 4679–4684.

[38] S. Paul, G. Boylan, S. Connolly, W. Marnane, and G. Lightbody, "A method for the blind separation of sources for use as the first stage of a neonatal seizure detection system," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, Philadelphia, PA, Mar. 2005, vol. 5, pp. 409–412, IEEE.

Stephen Faul was born on September 10, 1979, in Cork, Ireland. He received an honours degree in electrical and electronic engineering from University College Cork (UCC) in 2002, where he is currently pursuing the Ph.D. degree in the area of automated neonatal seizure detection.

His research interests include mathematical modeling of biological signals, nonlinear dynamic systems theory, classification networks, and independent component analysis. He is currently writing software for a 3-D imaging firm.

Gregor Gregorčič received the degree in electrical engineering and computer science from the University of Maribor, Slovenia, in 1998, and the Ph.D. degree in electrical engineering from the University College Cork (UCC), Cork, Ireland, in 2004.

At present, he is with the Control and Automation Development Department, Anstalt für Verbrennungskraftmaschinen List (AVL), Graz, Austria. Recently, his research focused on Gaussian processes as a nonparametric approach to modeling and control of nonlinear systems. His current research is concentrated on the development of advanced control strategies applied to complex automation systems for the automotive industry.

Geraldine Boylan received the M.Sc. degree in physiology and the Ph.D. degree in clinical medicine from University College London, London, U.K.

She was a Clinical Scientist in neonatal medicine at Kings College Hospital London from 1996 to 2001. She is currently a Lecturer of pediatrics with the Department of Pediatrics and Child health, University College Cork, Cork, Ireland. Much of her more recent work is of an interdisciplinary nature and aims to create a synergy between medicine and engineering by using the skills and techniques of engineering signal processing research to address important medical problems, such as automated seizure detection.

Dr. Boylan was recently awarded a Science Foundation Ireland (SFI) Principal Investigator Career Advancement Award.

William Marnane received the B.E. degree in electrical engineering from the National University of Ireland, Cork, in 1984, and the Ph.D. degree from the University of Oxford, Oxford, U.K., in 1989.

He has been a Senior Lecturer with the Department of Electrical and Electronic Engineering, National University of Ireland, Cork, since 1999. His research interests include signal processing of EEG signals and digital design for signal processing.

Gordon Lightbody received the M.Eng. degree (with distinction) and the Ph.D. degree in electrical and electronic engineering from Queen's University Belfast, Belfast, U.K., in 1989 and 1993, respectively.

After completing a one-year postdoctoral position funded by Du Pont, he was appointed by Queen's University as a Lecturer of modern control systems. At Queens, his research focused primarily on the application of intelligent control, modeling, and fault detection/diagnosis based on multivariate statistical techniques. In 1997, he was appointed to a lectureship in control engineering at the National University Ireland, Cork. His current research interests include nonparametric modeling, local model networks for the modeling and control of nonlinear dynamical processes, fuzzy/neural systems, and nonlinear and adaptive control systems.

Sean Connolly received the M.B., M.Ch., and BAO degrees from the National University of Ireland, Cork, in 1984, the MRCPI degree from the Royal College of Physicians of Ireland in 1988, and the M.D. degree from the National University of Ireland in 1994.

From 1989 to 1996, he worked at and trained in Clinical Neurophysiology, The Middlesex Hospital, London, U.K.; Newcastle General Hospital, U.K.; and the Massachusetts General Hospital, Boston. Since July 1996, he has been a Consultant in clinical neurophysiology at St. Vincent's University Hospital, Dublin, Ireland.

Dr. Connolly was elected Fellow of the Royal College of Physicians of Ireland in 1998. He is currently Dean of the Irish Institute of Clinical Neuroscience and also a member of the British Society of Clinical Neurophysiology, the Association of British Neurologists, the American Association of Neuromuscular and Electrodiagnostic Medicine, and the American Academy of Neurology.